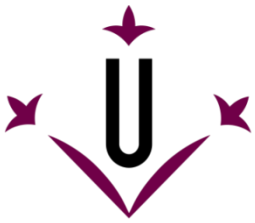


A probabilistic Poisson-based model to detect PRRSV recirculation using sow production records

L. Fraile[†1*], N. Fernández[†2], R.N. Pena[1], S. Balasch[3], G. Castellà[2], P. Puig[4], J. Estany[1], J. Valls[2,4]

[1]Department of Animal Science, University of Lleida – Agrotecnio Center, Lleida, Spain.

[2]Biostatistics and Epidemiology Unit, Biomedical Research Institute of Lleida (IRB Lleida), Lleida, Spain. º

[3]Department of Applied Statistics and Operational Research, Universitat Politècnica de Valencia, Valencia, Spain

[4]Department of Mathematics, Universitat Autònoma de Barcelona, Barcelona, Spain.

[†] These authors equally contributed to the paper

[*] Corresponding author: lorenzo.fraile@ca.udl.cat

Abstract

Porcine reproductive and respiratory syndrome (PRRS) is a viral disease associated with a decrease in the number of born alive piglets (NBA) and an increase in the number of lost piglets (NLP) per farrowing. Under practical conditions, it is critical to assess whether a farm is suffering PRRSV recirculation in the sow herd as soon as possible. The aim of this research work was to develop a new method to detect potential PRRSV recirculation in sow production farms. Sow reproductive performance records from one farm (farm T) were used to set up the method and records from ten additional farms (farms V1 to V10) were used for validation. A conditional Poisson model of NLP on NBA was proposed to fit the data. A three-step procedure was implemented to detect potential PRRSV recirculation: (i) computation of the maximum-likelihood estimates of the expected values of NBA and NLP in a PRRSV non-recirculating scenario; (ii) calculation, for each farrowing, of the p-value associated with the probability of jointly observing deviations towards decreased NBA and increased NLP. The detection of a potential PRRSV recirculation was based on (iii) the combined p-value resulting from weighing the p-values of the last N farrowings by the chi-square-inverse method. In order to gain specificity, a displacement on the expected non-recirculating NBA and NLP values was used for tuning purposes. With this approach, two PRRSV circulating periods were detected in farm T, which were confirmed with standard laboratorial diagnostic techniques. The method was subsequently validated in farms V1 to V10, where ten PRRSV-recirculating time episodes had been diagnosed. The method proposed here was able to detect the ten PRRSV recirculations using a relatively small set of contiguous farrowings, with only two mismatched weeks, one as a false negative, in farm V1, and one as a false positive, in farm V4. It is concluded that a conditional Poisson-based model of NLP on NBA can be a useful tool for routinely detecting PRRSV recirculation in sow herds.

Key words: PRRSV circulation, conditional Poisson distribution, detection

Introduction

Porcine reproductive and respiratory syndrome (PRRS) is the viral disease that causes the highest economic losses to the swine industry either in North America, Europe or Asia (Rowland et al., 2012). The major impact of PRRSV is reproductive failure in sows, which present a higher rate of abortion, as well as farrowings with less number of piglets born alive (NBA) and more number of lost piglets (NLP) (Lunney et al., 2010). The prevalence of PRRSV is high in regions under intensive pig farming systems (Fraile et al., 2010), so that both naïve and PRRSV positive farms in these areas can expect to suffer frequent PRRSV outbreaks. The diagnosis of the disease is commonly based on clinical suspicion followed by laboratory confirmation with standard diagnostics procedures such as RT-PCR and antibody detection with an ELISA technique. Clinical suspicion becomes evident if a sudden increase in abortion rate and/or a relevant decrease in NBA and/or an increase of NLP is observed. However, this is often not the case, particularly if sows are infected with a PRRSV strain of low virulence or they have developed previous immunity, as happens in PRRSV positive farms. In this situation, it would be highly beneficial to have reliable real-time alert triggers of PRRSV recirculation previously to laboratory confirmation (Linhares et al., 2016).

Lewis et al. (2009) suggest detecting a PRRSV outbreak using the raw 30-day rolling average for mummified pigs. Thus, deviations above the 99% confidence threshold of the average were considered to be part of an outbreak. To improve sensitivity, Rashidi et al. (2014) proposed the use of linear estimates of the herd-year-week effects for a single reproduction trait such as NBA or NLP to discriminate between healthy and disease phases. In this way, two or more consecutive weeks showing strong deviations from the overall herd-year-week mean identify a PRRSV

78 outbreak. With this method, sensitivity increased to 78%, but less severe outbreaks were
79 poorly detected (Mathur et al., 2014). Other methods extended this approach to include
80 more than one trait in the discrimination criterion (Mathur et al., 2014; Scanlan et al.,
81 2019). In particular, Mathur et al. (2014) based the detection of disease outbreaks on a
82 challenge load index combining data on NBA, NLP and number of weaned piglets of
83 individual sows. This method was able to detect outbreaks with sensitivity up to 93%.
84 Similar sensitivity values (>85.7%) were obtained by the application of statistical
85 process control techniques using exponentially weighted moving average on number of
86 abortions, pre-weaning mortality rate and prenatal losses (Silva et al., 2017). However,
87 all of these multi-trait methods were based on *ad hoc* procedures that ignore the joint
88 distribution from which the simultaneous outcome is sampled.
89 The aim of this research work is to develop a simple and robust method to indicate
90 PRRSV recirculation by modelling NBA and NLP simultaneously using a discrete joint
91 density based on Poisson distributions. The method is sought to outline, in real time and
92 at a large scale, deviances from the expected behavior of these two PRRSV-tagging
93 traits that trigger for a suspicion of PRRSV recirculation.
94
95 **2. Materials and Methods**
96 *2.1. Data*
97 Reproduction performance records from eleven sow production farms were used (Table
98 1). The model was trained using data from one farm (farm T) and subsequently
99 validated with the data from other ten farms (farms V1 to V10). Farm T was chosen to
100 set up the model because it was initially PRRSV negative and then had two diagnosed
101 outbreaks throughout a 9-year period. In each farrowing, the date of the farrowing and
102 NBA and NLP were recorded for each sow (epidemiological unit). Both NBA and NLP

were determined after the farrowing and NLP was calculated as the sum of stillborn and mummified piglets because these two traits can be easily misdiagnosed in commercial recording schemes. In farm T, data from 18,442 farrowings (from 4,519 sows) were collected from August 2006 to January 2015. Mean NBA and NLP were 12.99 and 1.77 piglets, respectively. The ten farms used for validation belonged to independent pig integration companies and used different genetic lines, nutrition and management procedures. Pigs were not vaccinated for PRRSV in any of these farms throughout the observation period. The PRRSV health status of each farm was monitored throughout the trial following the recommendations in Holtkamp et al. (2011) based on the exposure status of newborn piglets (less than two-day-old piglets). Serum samples of sixty piglets were tested for PRRSV weekly by direct detection of the virus through a quantitative reverse transcriptase PCR (qRT-PCR) assay. This sampling strategy allows detecting the presence of the virus with prevalence equal or higher than 5% and with a confidence level of 95%. Sampled piglets were randomly chosen unless weak newborns were present. Weak piglets were intentionally selected to increase the sensitivity to detect PRRSV. Then, viremia was measured using a semi-quantitative TaqMan PCR assay for PRRSV RNA in pooled samples (three samples per pool) as previously described (Abella et al., 2016). The duration of PRRSV recirculation in a given sow herd was established as the number of weeks in which PRRSV was detected in at least one of the pools. Ten PRRSV recirculating periods were confirmed in seven farms (farms V1, V5 and V10 had two outbreaks) while no PRRSV outbreaks were detected in three of them (V2, V6 and V9). The mean duration of PRRSV recirculating periods was 14 weeks (ranging from 6 to 17 weeks) and, in total, validation farms underwent 135 weeks under PRRSV positive diagnosis (Table 1).

127

128 *2.2. Statistical procedures*

129 *2.2.1 Statistical model*

130 Let X and Y be two dependent random count variables and $P(X = x, Y = y)$ their
131 discrete joint density. For the purpose of this work, X refers to NBA and Y to NLP,
132 although they can be any other PRRSV-tagging trait. In broader terms, $P(X = x, Y = y)$
133 should be flexible enough to admit any correlation structure between x and y. Since a
134 density function like this is not readily available, we used a bivariate conditional
135 Poisson model (Berkhout and Plug, 2004), in which the assumptions are made on the
136 marginal and conditional distributions in lieu of $P(X= x, Y=y)$. According to conditional
137 probability theory, $P(X = x, Y = y)$ can be written as the product of a marginal ($P_1(X =$
138 $x)$) and the conditional distribution ($P_2(Y = y \mid x)$). It is then assumed that P_1 and P_2 are
139 both Poisson distributions and hence

140

141 $$P(X = x, Y = y \mid \lambda_1, \lambda_2) = P_2(Y = y \mid x)\, P_1(X = x) = \frac{e^{-\lambda_2}\lambda_2^y}{y!}\ \frac{e^{-\lambda_1}\lambda_1^x}{x!} \qquad [1]$$

142

143 where x and y are the realized values of the random variables X and Y, respectively, with
144 values 0, 1, 2, …, and λ_1 and λ_2 are the parameters of the distributions. To allow for
145 interdependence between X and Y, λ_2 was set to depend on the realized value x of X.
146 Thus, $\lambda_2 = e^{k+\Theta x}$, where k is an intercept and Θ the regression coefficient. In this setting,
147 the expectation (E[Y]) and the variance (V[Y]) of the marginal distribution of Y are
148 $E[Y] = k\, e^{\lambda_1 (\exp(\Theta)-1)}$ and $V[Y] = E[Y]+E[Y]^2 (e^{\lambda 1(\exp(\Theta)-1)^2}-1)$. The correlation ($r_{xy}$)
149 between X and Y is as follows in equation 2:

150

151 $$r_{xy} = \frac{\lambda_1 \mathrm{E[y]}(\exp(\alpha) - 1)}{\sqrt{\lambda_1 \mathrm{E[y]}(1 + \mathrm{E[y]}\left(e^{\lambda_1(\exp(\alpha)-1)^2} - 1\right))}} \quad [2]$$

152

153 Note that if ϴ=0, r_{xy} =0 and the model reduces to a double Poisson distribution

154 (i.e. to the product of two independent Poisson distributions). In equation [1] Y is

155 conditioned to X, but it could have carried out inversely, leading to a different solution.

156 The choice of which permutation to use depends on the underlying causality structure.

157 Here, although both permutations have been assessed, we consider NLP (*Y*) conditioned

158 to NBA (*X*) as the reference situation.

159

160 *2.2.2. PRRSV recirculation detection*

161 Evidence of suspicious PRRSV recirculation is taken after a three-step process

162 (Figure 1). In step 1, the maximum likelihood estimates of λ_1, ϴ and k in equation [1]

163 and [2] under a PRRSV non-recirculating situation are obtained and denoted λ_1^0 and λ_2^0,

164 respectively. Here, these estimates were obtained using a subset of data from farm T

165 produced during a PRRSV negative period, as confirmed by gold standard diagnostic

166 procedures (RT-PCR negative). The values of λ_1^0 and λ_2^0 are used for hypothesis testing (

167 H_o:$\lambda_1 \geq \lambda_1^0$;$\lambda_2 \leq \lambda_2^0$), where the null hypothesis (H_0) assumes that (i) the expected

168 value for *X* (NBA) is at minimum the one observed in the non-circulating scenario and

169 (ii) the expected value for *Y* (NLP) is at maximum the one observed in the non-

170 circulating scenario. Indeed, this null hypothesis is testing whether the expected values

171 of NBA (λ_1) are equal or higher than a given threshold (λ_1^0) and the expected values of

172 NLP (λ_2) are equal or lower than another given threshold (λ_2^0). This hypothesis

173 formulation should be interpreted as a bivariate one-sided contrast against the non-

174 circulating situation.

In step 2, for each farrowing, the joint probability of X being lower than λ_1^0 and Y being greater than λ_2^0 is calculated. To gain in specificity, H_0 can be refined with the addition of two displacement parameters (ϕ_1 and ϕ_2) in the direction of the alternative hypothesis (lower X and greater Y), such that $H'_o{:}\lambda_1 \geq \lambda_1^0 - \phi_1; \lambda_2 \leq \lambda_2^0 + \phi_2$. Note that the greater ϕ_1 and ϕ_2, the more conservative the discriminant p-value (see definition in step 3). Sensible values for ϕ_1 and ϕ_2 were obtained after tuning the model with data collected from known circulating and non-circulating periods in farm T. Since X and Y are correlated, the value of ϕ_2 can be linearly interpolated from ϕ_1. With this simplification, a single displacement parameter (ϕ_1) suffices to tune the model for PRRSV circulating detection. Thus, ϕ_1 is a tuning parameter that needs to be assessed using a range of possible values. With this rationale, different significance boundaries in the $X \times Y$ plane can therefore be established. The final value of ϕ_1 should be interpreted as the maximum accepted decrease in the expected values of X and increase in the expected values of Y. In our analysis, we investigated values of ϕ_1 ranging from 2 to 4 units (piglets), which corresponded to ϕ_2 values ranging from 0.68 to 1.37. Obviously, this is a data-dependent issue and, naturally, ϕ_1 and ϕ_2 could also be determined by the user using any other criterion.

The step 3 develops the assessment criterion for potential PRRSV recirculation. To this aim, a combined p-value is calculated by weighing the p-values associated with the last N farrowings using the chi-square-inverse method (Fisher, 1950; Pearson, 1933). A combined p-value of less than 10^{-5} was considered to be statistically significant for suspicious recirculation.

The statistical software R was used for the analyses (RC Team, 2016). In particular, the 'bivpois' package is used for maximum likelihood estimation of the parameters of bivariate Poisson regression models (Karlis and Meligkotsidou, 2005).

200

201 *2.3. Validation*

202 Step 2 and 3 from the method was run on data from farms V1 to V10 using the
203 parameters set up with the training set. All detected outbreaks were validated against
204 analytical diagnosis as described above. Each week all farms used for validation
205 purposes were classified as either positive or negative and the resulting diagnosis was
206 matched with the predicted by the model. Similarly to other diagnostic assays, the
207 sensitivity and the specificity of the proposed method for detecting PRRSV outbreaks
208 were calculated using free epidemiological software (https://epitools.ausvet.com.au). As
209 for any diagnostic procedure, the sensitivity and specificity of the method should be
210 close to 100% and not less than 90%.

211

212 **3. Results**

213 Monthly evolution of NBA and NLP for the T farm showed evidences of
214 potential PRRSV circulation in 2008 (Figure 2A), where NBA and NLP decreased and
215 increased, respectively. No other clear PRRSV-circulating period was apparently
216 observed by visual inspection.

217 Data distribution features for NBA and NLP support that the Poisson assumption
218 can be a useful approach. In Figure 2B the observed and the expected probability
219 distribution of NBA and NLP are represented assuming that both traits follow univariate
220 Poisson distributions. As expected under Poisson assumptions, the sample mean and
221 variance estimated for NBA were very close to the real value. However, they were not
222 for NLP, where the inspection of the data revealed, in addition to a highly asymmetric
223 and zero-inflated distribution, over-dispersion. This is indicative that single Poisson
224 distributions might not be optimal for modeling NLP. On the other hand, in agreement

225 with previous results in the literature (Nielsen et al., 2014), we found that NBA and
226 NLP were negatively correlated (ρ= -0.24, 95% CI -0.26,-0.23, p<0.00001, Table 2).
227 The conditional Poisson model intends to undertake these two limitations. In accordance
228 with the underlying causality structure, NLP conditioned to NBA behaved better than
229 vice versa, which involves assuming λ_2 as the expectation of NLP conditioned to NBA.
230 Model estimates under this permutation were closer to sample estimates and over-
231 dispersion of NLP was sensibly reduced. The estimated correlation between NBA and
232 NLP under this model was -0.31. In Figures 3A and 3B, the observed and the expected
233 bivariate conditional joint probability using the conditional Poisson model with NLP
234 conditioned to NBA are shown, suggesting that the modelling approach captures the
235 structure of the data, particularly the negative correlation

236 In farm T, the PRRSV-circulating phase produced farrowings where NBA and
237 NLP had 3.4 piglets less and 1.2 piglets more as compared to the non-circulating phase
238 (Table 3). Moreover, the variance of both NBA and NLP increased under a PRRSV-
239 circulating phase, with values ranging from 1.5 times higher (NBA) to almost 2.5 higher
240 (NLP) than in a non-circulating phase. Conversely, the non-circulating situation
241 evidenced under-dispersion for NBA.

242 A visual representation of the combined p-values associated with the null
243 hypothesis of PRRSV non-circulating in farm T over time is shown in Figure 4. The
244 combined p-value at each time point was obtained weighting the p-values of the last N
245 farrowings. Individual p-values were calculated by conditioning the Poisson model of
246 NLP to NBA with different displacement parameters (ϕ_1=0 to 4). Specificity in
247 detection of PRRSV outbreaks was very poor for ϕ_1=0 but increased at higher values.
248 In our training data set, the optimal displacement stands at around ϕ_1=3 (□$_2$=1.03).
249 With these displacement values, not only the visually evident PRRSV circulation

250 occurred in 2008 was detected (Figure 2A) but also a new one apparently occurring in
251 2012. This second PRRSV-circulating period, which in fact was confirmed by qRT-
252 PCR, was an unexpected finding not possible to infer from a simple inspection of the
253 data (Figure 2A). Moreover, two additional unnoticed rebound outbreaks just after each
254 of the two main PRRSV-circulating periods were identified. As can be observed by
255 exploring the monthly average of NBA and NLP over these two periods, the first
256 outbreak prompted a relevant change in both NBA and NLP, while the second one only
257 a small increase in NLP, which made it more difficult to unravel.

258 The methodology proposed here was able to detect the ten PRRSV recirculation
259 time episodes occurred in the farms used for validation, with only one week, in farm
260 V1, being a false negative, and one week, in farm V4, being a false positive. The
261 sensitivity and specificity of the method was 99.3 (CI 98-100) and 99.99 % (CI 99.90-
262 100), respectively. Figure 5 and Figure 6 plot, respectively, the time-series of final p-
263 values (expressed as –log P-values) in farm V5, which had two confirmed PRRSV-
264 circulating periods, and in farm V9, which had no confirmed PRRSV outbreaks. In farm
265 V5, –log p-values were as high as 5 to 15 during the PRRSV-circulating period, while
266 in farm V9 they were always lower than 4.

267

268 **4. Discussion**

269 A new statistical approach is described to detect potential PRRSV-circulating
270 periods using the joint evolution of NBA and NLP in a particular farm over time. The
271 approach used was based on a bivariate conditional Poisson model for NBA and NLP
272 (Berkhout and Plug, 2004) instead of the bivariate density function, which is only
273 available for positive correlated variables. As an alternative, we could have used the
274 number of total piglets born (NBA +NLP) and NLP as discriminant variables, given that

they are positively correlated. However, we preferred to develop a conditional model that enables a flexible configuration of the input variables as far as their correlation structure. Under this setting, the density function is not unique and therefore one of the two alternative conditional permutations must be selected. In practical terms, this is not a problem as long as the causal variable is known, as here, or ϴ is not very high, something that is also expected to happen between litter size and mortality. As compared to other Poisson models (such as the double Poisson, bivariate Poisson or diagonal-inflated bivariate Poisson), the conditional Poisson of NLP on NBA was the one that best fit to the data (data not shown).

The methodology developed using this model can efficiently detect PRRSV-circulating periods not only when overt changes in NBA and NLP took place but also in case of subtle changes in either of the two. The set up used here has proved powerful enough for the farms used a case study. The method, however, still allows for further improving and tuning in order to take into account specific needs of more heterogeneous situations. Thus, the model could be extended to account for different effects (β), either fixed (such as the parity) or random (such as the sow), by modeling λ_1 and λ_2 as a linear function of them (i.e. $\lambda_i = e^{x'\beta}$, where x' is their incidence vector for a given farrowing). There are three parameters that can be adjusted to tune the model for PRRSV detection: i) the displacement parameters ($\square_1$ and $\square_2$); (ii) the number of parities used in the moving average (N); and (iii) the threshold for suspicious PRRSV circulation (combined p-value). We fixed these parameters to values that appeared to be optimal in our conditions, but obviously they can be tuned after training to deal with more heterogeneous situations. Similar tuning issues have been raised by others authors (Mathur et al., 2014).

299 The diagnosis of reproductive disorders is challenging because it can be
300 influenced by a diverse number of infectious and non-infectious factors. Thus, in
301 general practice, a correct diagnosis is based on a combination of performance records,
302 clinical observation, pathological examination and laboratory testing (Segalés et al.,
303 2013). Pathological and laboratory testing involves specialized equipment and personnel
304 and thus an extra cost for the farmer. Contrarily, performance records are available at no
305 additional cost to producers, since almost all of them have and use software to monitor
306 their farm's reproductive outcome. A limiting feature of these pig management systems
307 is that they do not analyse the data in real-time and thus they cannot give alarm signals
308 if warning criteria are out of the expected value. We have shown that monitoring NBA
309 and NLP can be useful criteria to detect a PRRSV infection, given that the progression
310 of the disease usually entails a relevant decrease in NBA and increase in NLP (Lunney
311 et al., 2010). Nevertheless, variations in NBA and NLP are non-specific and so they can
312 be due to many other infectious or non-infectious causes. As a result, the statistical tool
313 developed here does not preclude the need for laboratorial diagnosis confirmation but
314 provides a new element for optimal decision-making in terms of sampling for
315 pathological and laboratory examinations.
316 Other traits than NBA and NLP have been used as a criterion to detect PRRSV
317 outbreaks. Amongst them, owing to their direct role in the disease, the number of
318 mummified piglets (Lewis et al., 2009; Mathur et al., 2014) and the number of abortions
319 (Silva et al., 2017; Scanlan et al., 2019) are the two mostly considered. The procedure
320 developed allows the users to individualize the mortality trait to monitor, whether NLP
321 or the number of mummified. Here we chose to use NLP because it includes the number
322 of both stillborn and mummified piglets, two traits that can be easily misdiagnosed in
323 commercial recording schemes. So far we have not worked specifically with the number

of abortions, which can be overlooked under current requirements for animal welfare. With many sows per pen, daily monitoring of individual gestation status becomes more burdensome.

The method has been designed to improve specificity of detection as farmers dislike false alarms. For this reason, the diagnostic performance of the method must be interpreted with caution until confirmed in more farms. The parameters of the model can be estimated as described above from own data or use the ones obtained here as default values. However, ad hoc parameters can be derived for each case including different genetic types or management practices. A potential limitation of the method could be the lack of enough statistical power in small farms, where the small number of farrowings per week may lead to delayed detection of outbreaks. Moreover, it should be remarked that the method has only been validated in non-vaccinated farms, so as to define recirculation periods without the need of sequencing to distinguish between wild-type and MLV PRRSV. Thus, it would be also convenient to test the method in PRRSV vaccinated farms. In conclusion, a new statistical approach based on time series of data on litter size and mortality has been developed to detect with high sensitivity and specificity PRRSV-circulating periods in sow farms.

Conflict of interest statement

The authors declare no conflict of interests that could inappropriately influence or bias the content of the paper.

Acknowledgments

This study was carried out with the support of many pig producing companies that shared the data with authors. Moreover, this study was partially funded by the FEDER project with reference COMRDI16-1-0035-03.

References

Abella, G., Pena, R.N., Nogareda, C., Armengol, R., Vidal, A., Moradell, L., Tarancon, V., Novell, E., Estany, J., Fraile, L. 2016. A WUR SNP is associated with European Porcine Reproductive and Respiratory Virus Syndrome resistance and growth performance in pigs. Res. Vet. Sci. 104, 117-122.

Berkhout, P. and Plug, E. 2004. A Bivariate Poisson Count Data Model Using Conditional Probabilities. Stat Neerl. 58(3), 349-364.

Fisher, R.A.1950. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh;.

Fraile L, Alegre A, López-Jiménez R, Nofrarías M, Segalés J. 2010. Risk factors associated with pleuritis and cranio-ventral pulmonary consolidation in slaughter-aged pigs. Vet J. 184(3), 326-33.

Holtkamp, D. J., Polson, D. D., Torremorell, M., Morrison, B., Classen, D. M., Becton, L., Zimmerman, J. 2011. Terminology for classifying swine herds by porcine reproductive and respiratory syndrome virus status. Journal of Swine Health and Production. 19(1), 44-56.

Karlis, D., and Meligkotsidou, L.2005. Multivariate Poisson regression with covariance structure. Stat Comput. 15(4), 255-265.

Lewis, C.R.G., Torremorell, M., Galina-Pantoja, L., Bishop, S.C. 2009. Genetic parameters for performance traits in commercial sows estimated before and after an outbreak of porcine reproductive and respiratory syndrome. J. Anim. Sci. 87(3), 876-884.

Linhares DC, Cano JP, Torremorell M, Morrison RB. 2014.Comparison of time to PRRSv-stability and production losses between two exposure programs to control PRRSv in sow herds. Prev Vet Med. 116(1-2), 111-119.

Lunney, J.K., Benfield, D.A., Rowland, R,R. 2010. Porcine reproductive and respiratory syndrome virus: an update on an emerging and re-emerging viral disease of swine. Virus Res. 154(1-2), 1-6.

Mathur, P.K., Herrero-Medrano, J.M., Alexandri, P., Knol, E.F., ten Napel, J., Rashidi, H., Mulder, H.A. 2014 Estimating challenge load due to disease outbreaks and other

challenges using reproduction records of sows. Journal of Animal Science. 92(12), 5374-5381.

Nielsen, B., Su, G., Lund, M.S., Madsen, P. 2013 Selection for increased number of piglets at d 5 after farrowing has increased litter size and reduced piglet mortality. Journal of Animal Science. 91(6), 2575-2582.

Pearson, K. 1933. On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. Biometrika. 25(3-4), 379-410.

Rashidi, H., Mulder, H.A., Mathur, P., van Arendonk, J.A., Knol, E.F. 2014. Variation among sows in response to porcine reproductive and respiratory syndrome. J. Anim. Sci. 92(1), 95-105.

Rowland, R.R., Lunney, J., Dekkers, J. 2012. Control of porcine reproductive and respiratory syndrome (PRRS) through genetic improvements in disease resistance and tolerance. Front. Genet. 3, 260.

Scanlan C. L., Putz, A.M., 1, Gray, K.A., Serão, N.V.L. 2019. Genetic analysis of reproductive performance in sows during porcine reproductive and respiratory syndrome (PRRS) and porcine epidemic diarrhea (PED) outbreaks. J. Anim. Sci. Biotechnol. 10:22

Segalés J, Martínez J, Castellà J, Darwich L, Domingo M, Mateu E, Martín M, Sibila M. Handbook of laboratory diagnosis in Swine, 2013. Editorial servet.

Silva, G.S., Schwartz, M., Morrison, R.B., Linhares D.C.L. 2017. Monitoring breeding herd production data to detect PRRSV outbreaks. Prevent. Vet. Med. 148, 89-93.

Team RC. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2014. 2016.

Figure caption list

Figure 1. Chart flow describing the key steps of statistical method**.**

Figure 2. Data description of the training data set (farm T). (A) Monthly NBA and NLP mean through over all the period analyzed. Dotted vertical red lines represent a potential PRRSV-circulating situation, i.e. decreased piglets born alive (NBA) and increased lost piglets *(*NLP). Dotted horizontal lines represent the global mean in the whole period. (B) Observed (empirical) and expected (under a univariate Poisson distribution) probability functions.

Figure 3. (A) Observed (empirical) and expected (under a bivariate Poisson model conditioned to number of piglets born alive, NBA) joint probability function for all possible NBA × NLP (number of lost piglets) combinations. (B) Observed (empirical) and expected (under a bivariate Poisson model conditioned to NBA) probability function in PRRSV-circulating (outbreak) and non-circulating (non-outbreak) phases. Probabilities are represented as colors ranging from red (the highest) to yellow (the lowest).

Figure 4. Evolution over time of the P-value resulting from weighting the P-values associated with the last n farrowings by the chi-square-inverse method. Individual P-values were calculated under a bivariate Poisson model of the number of lost piglets (NLP) conditioned to the number of piglets born alive (NBA), and either not using or using displacement parameters (ϕ_1 and ϕ_2).

Figure 5. Time-series plot of the –log P-value resulting from weighting the P-values of the previous n farrowings, with displacement (ϕ_1=3 and ϕ_2=1.03) in farm V5. This farm had confirmed PRRSV circulation from February to June 2016.

Figure 6. Time-series plot of the –log P-value resulting from weighting the P-values of the previous n farrowings, with displacement (ϕ_1=3 and ϕ_2=1.03) in farm V9. This farm had no PRRSV-circulating periods.

Table 1. Descriptors of the farms used for training (T) and validation (farms V1 to V10).

Farm	No of sows	No of monitored weeks	Number of PRRSV recirculating periods	PRRSV* status at the beginning of the study	Number of weeks with a PRRSV positive diagnosis (RT-PCR)
T	1,200	468	2	Naïve	26
V1	1,000	520	2	Positive stable	30
V2	1,100	104	0	Naïve	0
V3	1,200	156	1	Positive stable	15
V4	800	104	1	Positive stable	13
V5	730	520	2	Positive stable	12
V6	1,500	208	0	Naïve	0
V7	3,000	156	1	Positive stable	17
V8	2,500	208	1	Positive stable	14
V9	560	104	0	Naïve	0
V10	690	296	2	Positive stable	34

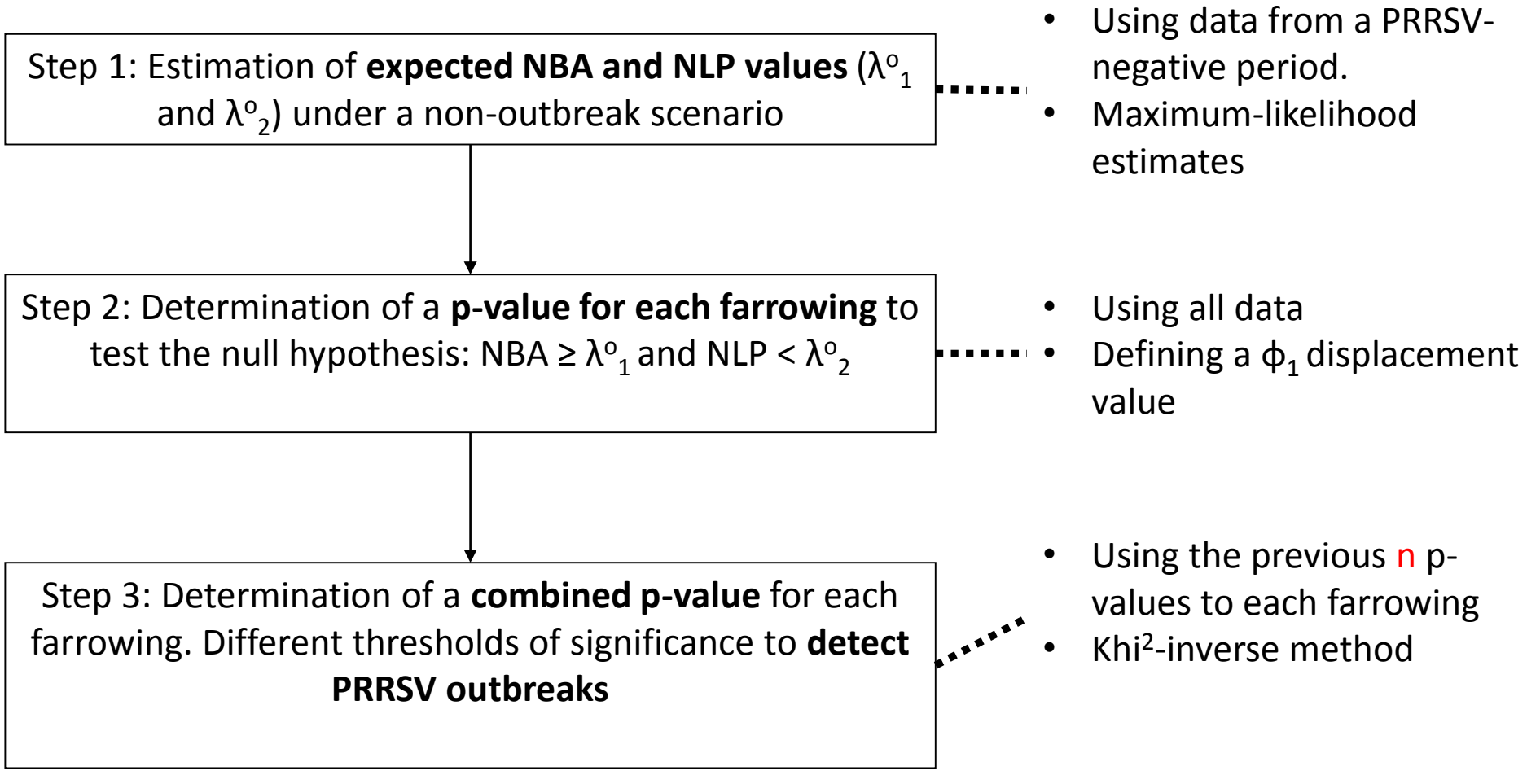
*The PRRSV health status of each farm was monitored throughout the trial following the recommendations by Holtkamp et al. (2011)

Table 2: Sample and model estimates of the expectation (E) and variance (V) for number of born alive (NBA) and number of lost piglets *(*NLP) per parity in farm T as well as and their covariance (cov) and correlation (r). Model estimates were obtained with a conditional Poisson model, either conditioning NLP to NBA or NBA to NLP.

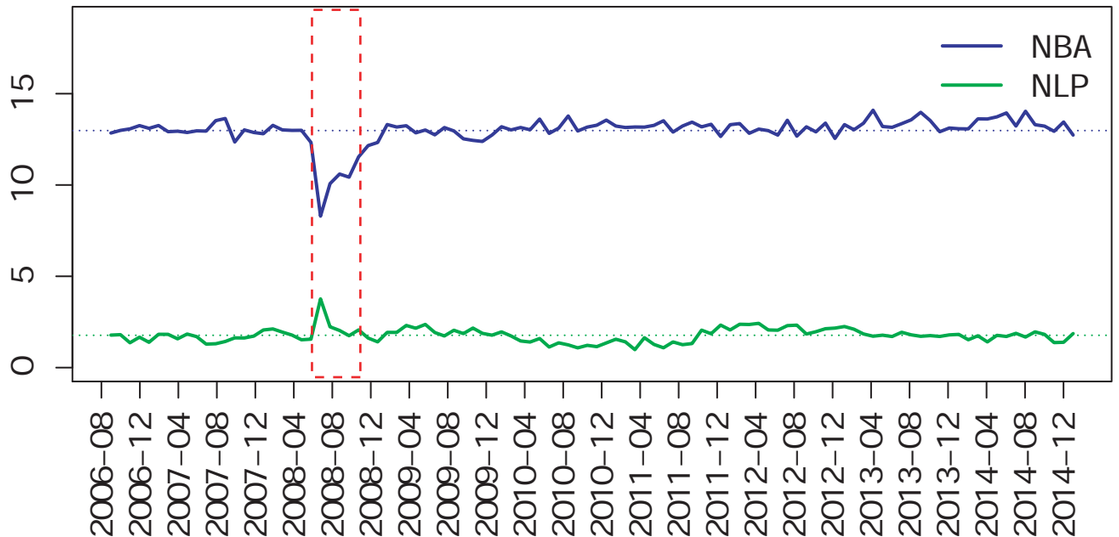
		Conditional Model estimates	
	Sample estimates	Conditioned to NBA	Conditioned to NLP
E[NBA]	12.99	12.98	12.97
V[NBA]	12.41	12.41	74.90
E[NLP]	1.77	1.77	1.77
V[NLP]	4.08	2.93	4.08
cov[NBA,NLP]	-1.74	-1.61	-0.80
r[NBA,NLP]	-0.24	-0.31	-0.16

Table 3: Sample and model estimates of the expectation (E) and variance (V) for number of born alive (NBA) and number of lost piglets *(*NLP) per parity in farm T as well as their covariance (cov) and correlation (r) in PRRSV-circulating and non-circulating phases. Model estimates were obtained with a conditional Poisson model where NLP was conditioned to NBA.

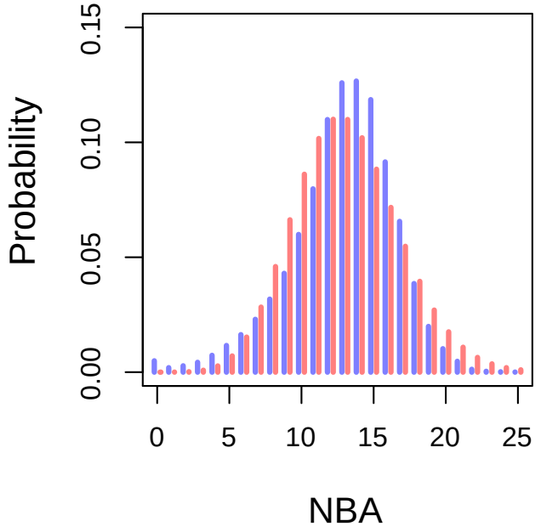
	Circulating phase		Non-circulating phase	
	Sample estimates	Model estimates	Sample estimates	Model estimates
E[NBA]	9.84	9.84	13.26	13.26
V[NBA]	17.70	17.70	11.70	11.70
E[NLP]	2.45	2.42	1.28	1.28
V[NLP]	5.50	4.59	2.58	1.89
cov[NBA,NLP]	-2.34	-1.22	-1.00	-1.02
r[NBA,NLP]	-0.23	-0.24	-0.18	-0.24
No. of farrowings	707		2,319	

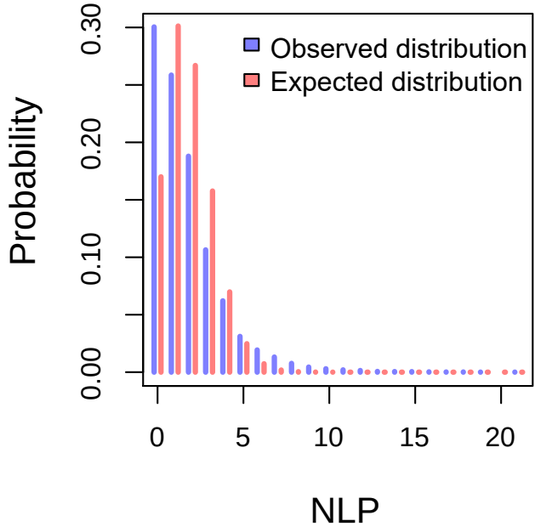

Step 1: Estimation of **expected NBA and NLP values** (λ^o_1 and λ^o_2) under a non-outbreak scenario
• Using data from a PRRSV-negative period.
• Maximum-likelihood estimates
Step 2: Determination of a **p-value for each farrowing** to test the null hypothesis: NBA ≥ λ^o_1 and NLP < λ^o_2
• Using all data
• Defining a ϕ_1 displacement value
Step 3: Determination of a **combined p-value** for each farrowing. Different thresholds of significance to **detect PRRSV outbreaks**
• Using the previous n p-values to each farrowing
• Khi²-inverse method

A

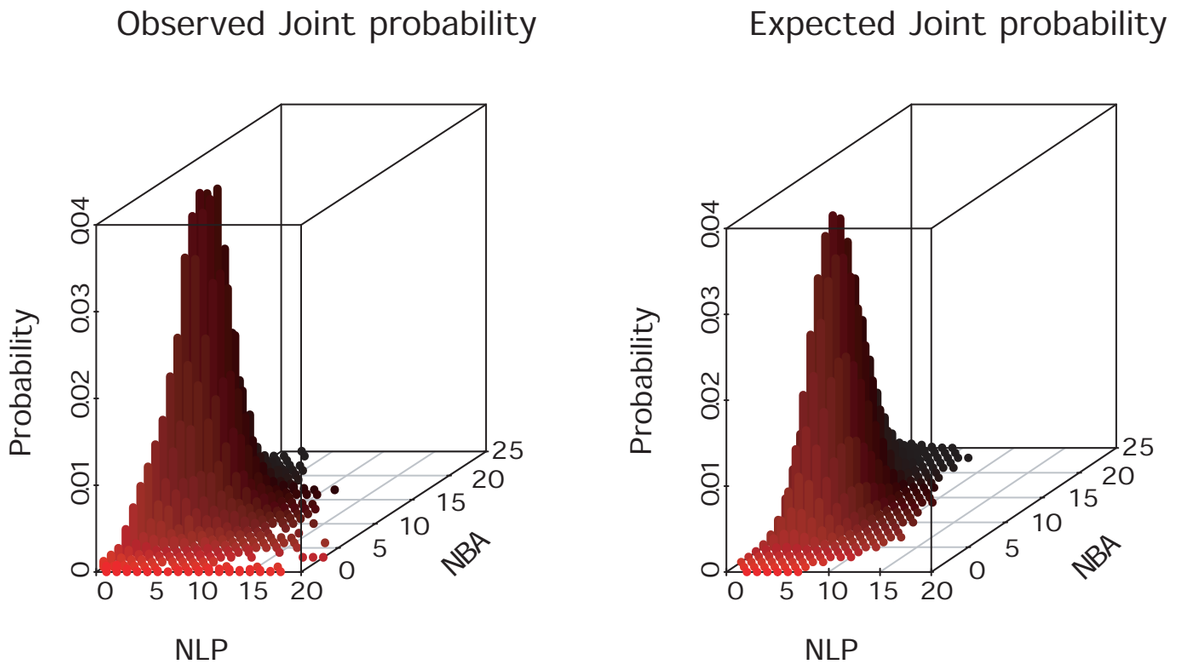

NBA
NLP
0
5
10
15
2006–08
2006–12
2007–04
2007–08
2007–12
2008–04
2008–08
2008–12
2009–04
2009–08
2009–12
2010–04
2010–08
2010–12
2011–04
2011–08
2011–12
2012–04
2012–08
2012–12
2013–04
2013–08
2013–12
2014–04
2014–08
2014–12

B

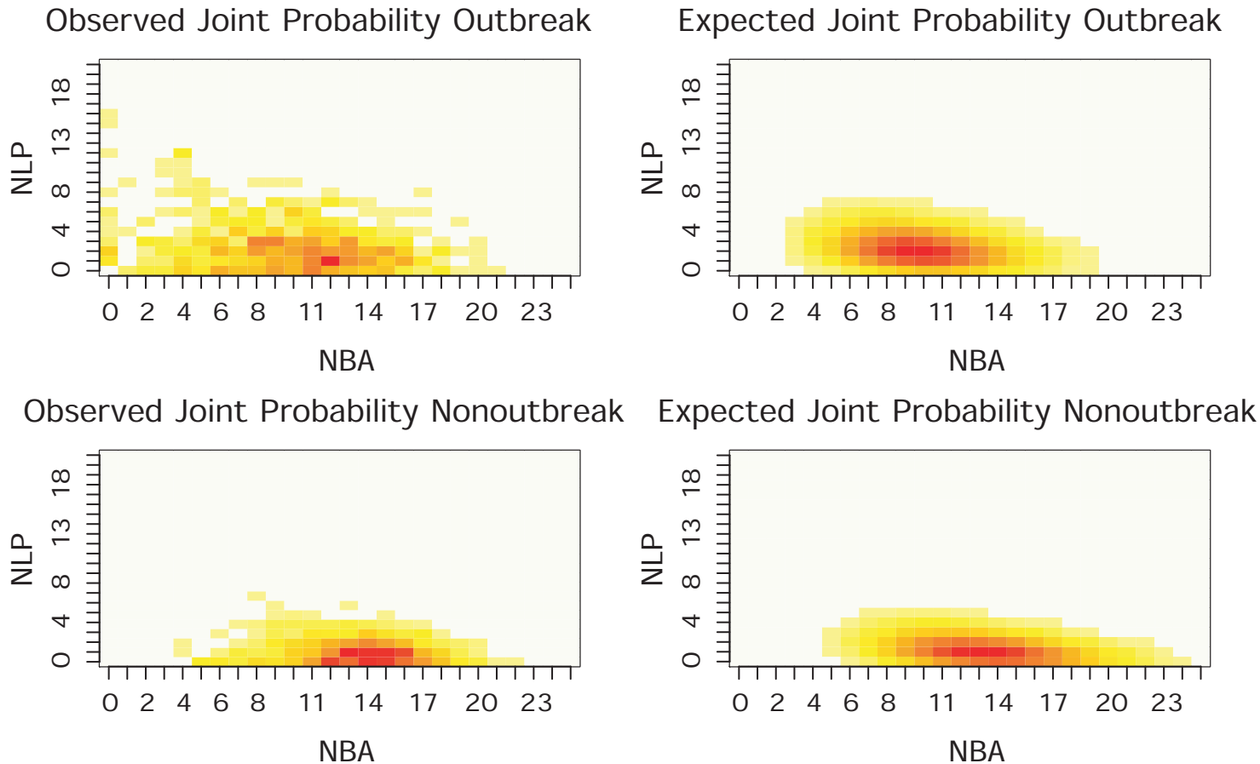

Probability
0.00
0.05
0.10
0.15
0
5
10
15
20
25
NBA

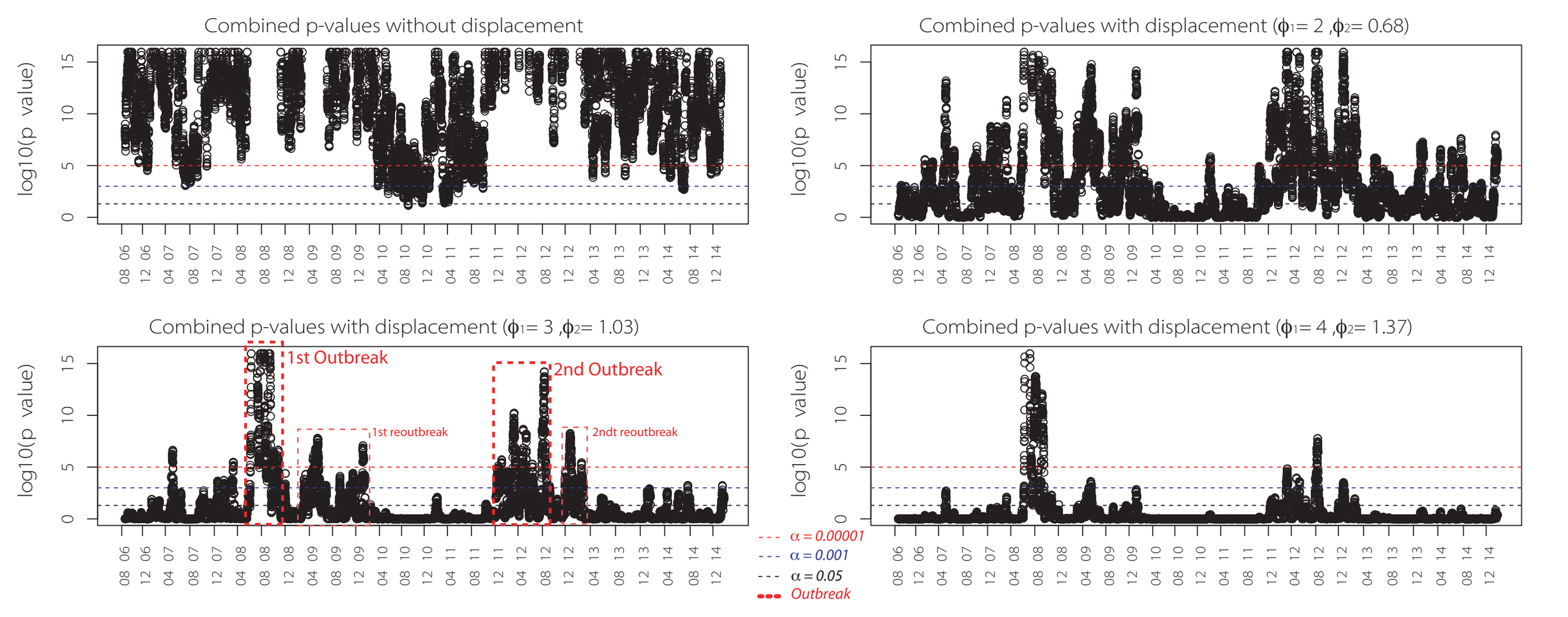

Probability
0.00
0.10
0.20
0.30
Observed distribution
Expected distribution
0
5
10
15
20
NLP

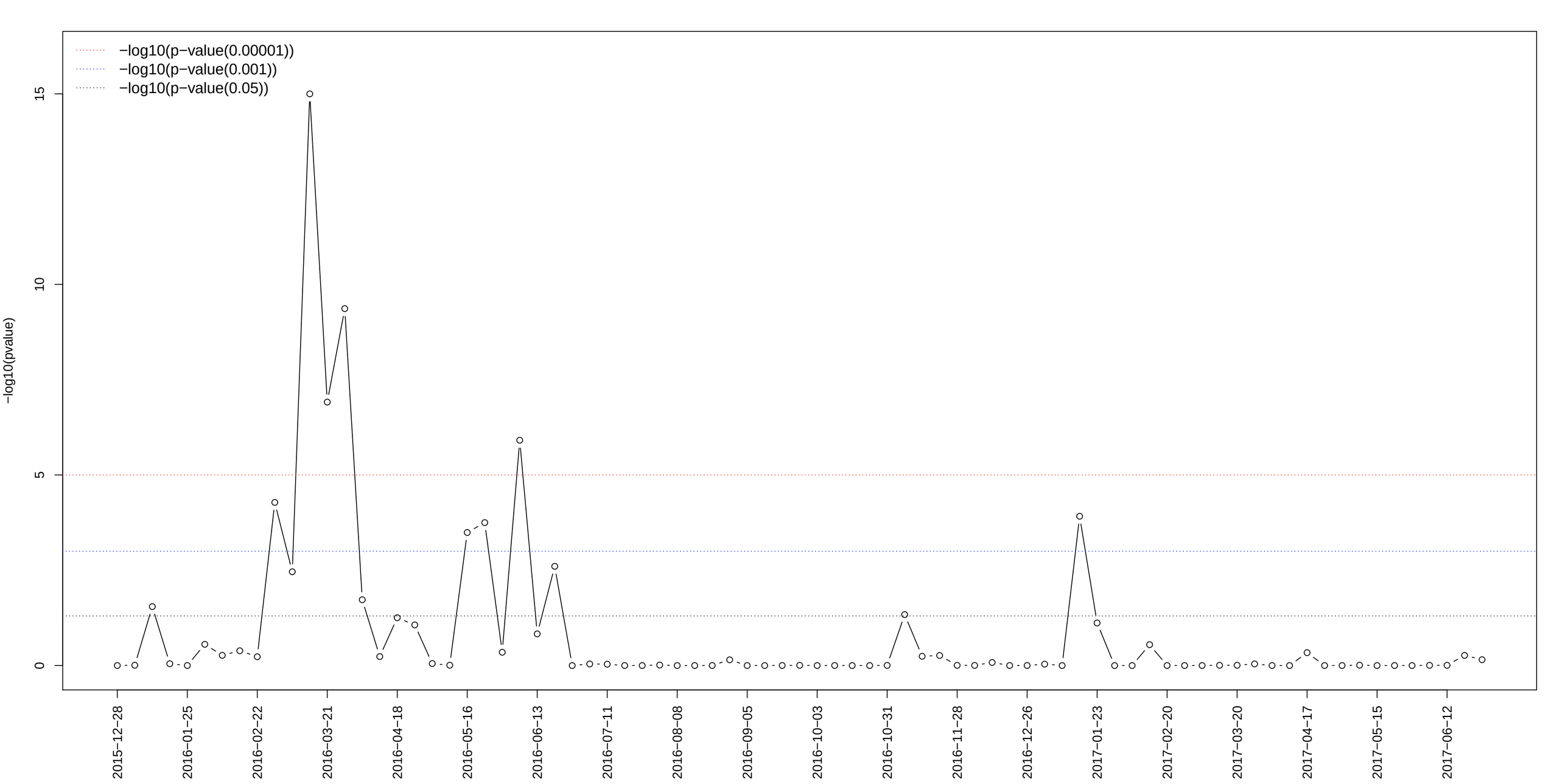
A

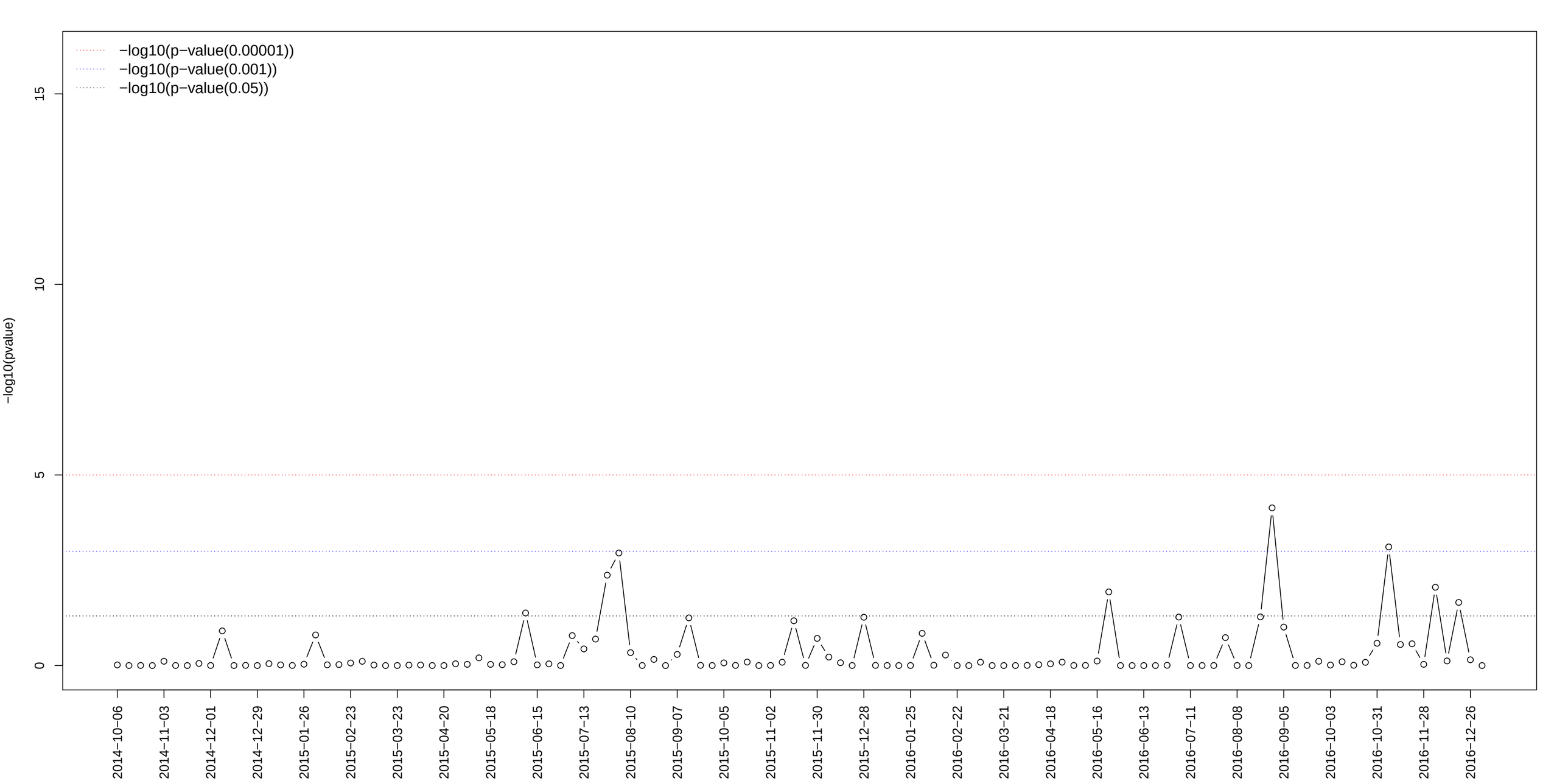

Observed Joint probability
Probability
0
0.01
0.02
0.03
0.04
0
5
10
15
20
NLP
0
5
10
15
20
25
NBA
Expected Joint probability
Probability
0
0.01
0.02
0.03
0.04
0
5
10
15
20
NLP
0
5
10
15
20
25
NBA

B


Observed Joint Probability Outbreak
NLP
0
4
8
13
18
0
2
4
6
8
11
14
17
20
23
NBA
Expected Joint Probability Outbreak
NLP
0
4
8
13
18
0
2
4
6
8
11
14
17
20
23
NBA
Observed Joint Probability Nonoutbreak
NLP
0
4
8
13
18
0
2
4
6
8
11
14
17
20
23
NBA
Expected Joint Probability Nonoutbreak
NLP
0
4
8
13
18
0
2
4
6
8
11
14
17
20
23
NBA


Combined p-values without displacement
Combined p-values with displacement (ϕ1= 2 ,ϕ2= 0.68)
Combined p-values with displacement (ϕ1= 3 ,ϕ2= 1.03)
Combined p-values with displacement (ϕ1= 4 ,ϕ2= 1.37)
log10(p value)
1st Outbreak
2nd Outbreak
1st reoutbreak
2ndt reoutbreak
α = 0.00001
α = 0.001
α = 0.05
Outbreak


−log10(p−value(0.00001))
−log10(p−value(0.001))
−log10(p−value(0.05))
−log10(pvalue)
0
5
10
15
2015−12−28
2016−01−25
2016−02−22
2016−03−21
2016−04−18
2016−05−16
2016−06−13
2016−07−11
2016−08−08
2016−09−05
2016−10−03
2016−10−31
2016−11−28
2016−12−26
2017−01−23
2017−02−20
2017−03−20
2017−04−17
2017−05−15
2017−06−12


−log10(p−value(0.00001))
−log10(p−value(0.001))
−log10(p−value(0.05))
−log10(pvalue)
0
5
10
15
2014−10−06
2014−11−03
2014−12−01
2014−12−29
2015−01−26
2015−02−23
2015−03−23
2015−04−20
2015−05−18
2015−06−15
2015−07−13
2015−08−10
2015−09−07
2015−10−05
2015−11−02
2015−11−30
2015−12−28
2016−01−25
2016−02−22
2016−03−21
2016−04−18
2016−05−16
2016−06−13
2016−07−11
2016−08−08
2016−09−05
2016−10−03
2016−10−31
2016−11−28
2016−12−26